Linear Programming Based Finite Blocklength Converses in Information Theory

Ankur. A. Kulkarni
Systems and Control Engineering
Indian Institute of Technology, Bombay, India
Email: kulkarni.ankur@iitb.ac.in

Sharu Theresa Jose
Systems and Control Engineering
Indian Institute of Technology, Bombay, India
Email: sharutheresa@iitb.ac.in

***Abstract*—A linear programming based framework is presented to derive finite blocklength converses for coding problems in information theory which is also extendable to network settings. In the point-to-point setting, the LP based framework recovers and in fact improves on almost all well-known finite blocklength converses for lossy joint source-channel coding, lossy source coding and channel coding. Moreover, the LP based framework is shown to be asymptotically tight for the averaged and compound channels under the maximum probability of error criterion. Further, for multiterminal Slepian-Wolf source coding problem, a systematic approach to synthesize new converses from considering point-to-point lossless source coding (with side-information at decoder) sub-problems is introduced. The method derives new finite blocklength converse for Slepian-Wolf coding which significantly improves on the converse of Miyake and Kanaya.**

I. Introduction

Fundamental to information theory is the problem of reliably transmitting information from a source to a destination by appropriately designing an encoder-decoder pair (called code); the reliability measured according to a *loss criterion*. Classical coding theorems of information theory consider transmission of information to be reliable when the loss criterion is minimized in the limit of large blocklengths, thereby requiring that an infinite blocklength (or infinite delay) is permitted in the system. However, most present day practical communication systems like the multimedia communication, cannot afford to have infinite delays in the system. Consequently, evaluating the optimal performance of the communication system for finite blocklengths is highly relevant.

Finite blocklength information theory pertains to evaluating the non-asymptotic or finite blocklength fundamental limit of communication. For example, in the channel coding problem with average probability of error as the loss criterion, this fundamental limit translates to finding the maximum rate of transmission across the channel for a fixed, finite blocklength such that the average probability of error does not exceed $\epsilon \in (0, 1)$. However, the exact evaluation of the finite blocklength fundamental limit is intractable in general for any coding problem, owing to its inherent non-convexity. Consequently, recent researches ([1],[2],[3],[4]) focus on obtaining best approximations of this finite blocklength fundamental limit of communication. Crucial to this is the derivation of 'good' *achievability* and *converses* that sandwich tightly the non-asymptotic limit. While finite blocklength *achievability* refers to an upper bound on the loss incurred by a code that can be guaranteed to exist for a given blocklength, the finite blocklength converse *converse* refers to a lower bound on the loss incurred by any code for a fixed blocklength.

Our interest in this work is on the finite blocklength converses or lower bounds on the loss criterion for fixed, finite blocklength. Although finite blocklength converses could be distilled from classical converses of coding theorems, these are rarely satisfyingly tight in the finite blocklength regime. Consequently, recent focus has been on obtaining tight finite blocklength converses. For the channel coding problem with average probability of error as the loss criterion, a metaconverse is derived by Polyanskiy, Poor and Verdú [1] employing *hypothesis testing*. For lossy joint source-channel coding and lossy source coding with probability of excess distortion as the loss criterion, tilted-information and hypothesis testing have been used to derive asymptotically tight converses by Kostina and Verdú [2], [3].

In the light of these results, our work is motivated by the following line of thoughts.

1) Even though many sharp converses have been derived for coding problems with different loss criteria, employing diverse techniques and approaches, what is perhaps unsatisfying is the absence of a *unified framework* to derive finite blocklength converses that applies to any *loss criteria* and that can conceptually situate all the known converses within itself.
2) While the derivation of finite blocklength converses in point-to-point information theory has been well-researched, extending many of these techniques to derive tight converses in the network setting has been proved to be quite challenging. Consequently, it would be an added bonus if the unified framework can be easily extended to *networked settings* to yield tight converses.

Derivation of such a unified framework has to effectively utilize the intrinsic non-convex geometry of the finite blocklength coding problem. Consequently, we resort to the tools and techniques in optimization to obtain a linear programming (LP) based framework for deriving finite blocklength converses for any loss criteria and is easily extendable to network settings. We explain the LP-based framework in the next section.

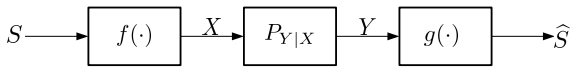

Fig. 1: Source-Channel Communication System

II. Linear Programming Based Framework

To explain the LP-based framework for deriving converses, consider the following general finite blocklength lossy joint source-channel coding problem SC,

$$\begin{array}{ll} \text{SC} & \min\limits_{f,g} \ \mathbb{E}[\kappa(S,X,Y,\widehat{S})] \\ \text{subject to} & \begin{array}{l} X = f(S), \\ \widehat{S} = g(Y). \end{array} \end{array}$$

Here $S, X, Y, \widehat{S}$ are random variables taking values in fixed spaces $\mathcal{S}, \mathcal{X}, \mathcal{Y}$ and $\widehat{\mathcal{S}}$ respectively and $\kappa : \mathcal{S}\times\mathcal{X}\times\mathcal{Y}\times\widehat{\mathcal{S}} \to \mathbb{R}$ is a given cost function. Since the spaces $\mathcal{S}, \mathcal{X}, \mathcal{Y}, \widehat{\mathcal{S}}$ are taken as fixed, the problem SC corresponds to a fixed blocklength setting (unit blocklength, if the alphabet is defined appropriately). Notice that these spaces could be spaces of strings of finite length. As before, S is the source signal distributed according to the probability distribution P_S, $f : \mathcal{S} \to \mathcal{X}$ represents the encoder which maps S to the encoded signal X (see Fig 1). The encoded signal is sent through a channel which given X produces an output signal Y according to a known channel law, denoted by $P_{Y|X}$, following which a decoder, $g : \mathcal{Y} \to \widehat{\mathcal{S}}$, maps the channel output signal to a destination signal $\widehat{S}$. Each pair f, g induces a joint distribution on $\mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \widehat{\mathcal{S}}$ and the expectation $\mathbb{E}$ is with respect to this joint distribution. Problem SC seeks to minimize the expectation of the loss function κ over all *codes*, i.e., over all encoder-decoder pairs (f, g).

The LP-based framework is motivated by an observation made in [5] that problems of the form of SC can be posed equivalently as a nonconvex optimization problem over joint probability distributions. Towards this, we introduce the following joint probability distribution,

$$Q(s,x,y,\widehat{s}) \equiv P_S(s)Q_{X|S}(x|s)P_{Y|X}(y|x)Q_{\widehat{S}|Y}(\widehat{s}|y), \quad (1)$$

such that $S \to X \to Y \to \widehat{\mathcal{S}}$ forms a Markov chain, $\sum_{s\in\mathcal{S},x\in\mathcal{X},y\in\mathcal{Y},\widehat{s}\in\widehat{\mathcal{S}}} Q(s,x,y,\widehat{s}) = 1$ and $Q(s,x,y,\widehat{s}) \geq 0$, for all $s \in \mathcal{S}, x \in \mathcal{X}, y \in \mathcal{Y}, \widehat{s} \in \widehat{\mathcal{S}}$. Here, $Q_{X|S}$ and $Q_{\widehat{S}|Y}$ are conditional probability distributions lying in the spaces $\mathcal{P}(\mathcal{X}|\mathcal{S})$ and $\mathcal{P}(\widehat{\mathcal{S}}|\mathcal{Y})$, respectively. $Q_{X|S}$ represents a *randomized encoder* and $Q_{\widehat{S}|Y}$ represents a *randomized decoder*. Together, $(Q_{X|S}, Q_{\widehat{S}|Y})$ constitute a *randomized code*.

Employing this joint probability distribution Q, SC can be equivalently posed as the following optimization problem over randomized codes,

$$\begin{array}{ll} \text{SC} & \min\limits_{Q,Q_{X|S},Q_{\widehat{S}|Y}} \sum\limits_{z} \kappa(z)Q(z) \\ & Q(z) \equiv P_S Q_{X|S} P_{Y|X} Q_{\widehat{S}|Y}(z), \\ & \sum_x Q_{X|S}(x|s) = 1 \quad \forall s \in \mathcal{S}, \\ \text{s.t} & \sum_{\widehat{s}} Q_{\widehat{S}|Y}(\widehat{s}|y) = 1 \quad \forall y \in \mathcal{Y}, \\ & Q_{X|S}(x|s) \geq 0 \quad \forall s \in \mathcal{S}, x \in \mathcal{X}, \\ & Q_{\widehat{S}|Y}(\widehat{s}|y) \geq 0 \quad \forall \widehat{s} \in \widehat{\mathcal{S}}, y \in \mathcal{Y}, \end{array}$$

where $\mathcal{Z} = \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \widehat{\mathcal{S}}$ and $z := (s, x, y, \widehat{s}) \in \mathcal{Z}$. We refer the readers to [6] for more details on the formulation.

Upon replacing Q in the objective function of SC with the RHS of first constraint, the resulting SC has a *bilinear objective function*- bilinearity arising due to the presence of product terms $Q_{X|S}(x|s)Q_{\widehat{S}|Y}(\widehat{s}|y)$. An important consequence of the bilinearity arising in SC is the resulting nonconvexity of the feasible region. A bilinear function is neither convex nor concave which makes evaluating the optimal value of SC intractable in general. Consequently, what one could hope for is to obtain upper and lower bounds that sandwich the $\mathrm{OPT}(\mathrm{SC})$ tightly.

What we are interested in is in obtaining tight converses or lower bounds on the optimal value of SC. A natural optimizer's approach to obtain converses would then be via a *convex relaxation* of the nonconvex optimization problem, which is a new optimization problem with a larger and convex feasible region containing the nonconvex feasible region of the original problem. Since the minimization is now done over a larger feasible region, the optimal solution of convex relaxation problem forms a lower bound on the optimal solution of the original nonconvex problem.

Towards obtaining a convex relaxation, we resort to the "lift-and-project" technique by Lovasz, Shrijver, Sharali, Adams and others in integer programming [7] and mechanistically derive a linear programming (LP) relaxation of the nonconvex optimization problem. Moreover, this can be systematically extended to derive LP relaxations for finite blocklength coding problems in multiterminal settings. The advantages of employing a linear programming relaxation, in addition to polynomial time solvability, is its duality. The duality of linear programming then ensures that the objective value of any feasible point of the dual of the LP relaxation is a lower bound on the optimal value of LP. Thus, the LP based framework to obtain finite blocklength converses asks to construct feasible points of the dual of the LP; with the constraints of the dual representing the sufficient conditions to establish an expression as a lower bound on the minimum expected cost.

Since the optimal solution of the original nonconvex problem and the LP relaxation lie on the extreme points of their respective feasible regions, the least one could expect from a 'good' LP relaxation is that it retains the extreme points of the original nonconvex problem in its set of extreme points. The LP relaxation that we derive has precisely this property. We present the LP relaxation and its dual in the next section.

A. LP Relaxation and Dual Program

To obtain a linear programming relaxation of problem SC, we resort to the following lift-and-project-like idea (see [8]). In this technique, problem SC is lifted to a higher dimensional space by introducing new variables $W(s,x,y,\widehat{s})$ to replace the product terms $Q_{X|S}(x|s)Q_{\widehat{S}|Y}(\widehat{s}|y)$, for all $s,x,y,\widehat{s}$. Subsequently, new valid inequalities are obtained employing these newly introduced variables in the lifted space. To obtain the valid inequalities, we adopt the following procedure. For each $s \in \mathcal{S}$, multiply the constraint $\sum_x Q_{X|S}(x|s) = 1$, with the variables $Q_{\widehat{S}|Y}(\widehat{s}|y)$ for all $\widehat{s} \in \widehat{\mathcal{S}}, y \in \mathcal{Y}$, and likewise for each $y \in \mathcal{Y}$ multiply the constraint $\sum_{\widehat{s}} Q_{\widehat{S}|Y}(\widehat{s}|y) = 1$, by $Q_{X|S}(x|s)$, for all $x \in \mathcal{X}, s \in \mathcal{S}$. We further obtain additional constraints by multiplying the variable bounds with each other, i.e. $(1-Q_{X|S}(x|s))(1-Q_{\widehat{S}|Y}(\widehat{s}|y)) \geq 0$, for all $s,x,y,\widehat{s}$ and $Q_{X|S}(x|s)Q_{\widehat{S}|Y}(\widehat{s}|y) \geq 0$, for all $s,x,y,\widehat{s}$. Hence, we obtain the following set of constraints, Subsequently, replace the bilinear product terms $Q_{X|S}(x|s)Q_{\widehat{S}|Y}(\widehat{s}|y)$ in the constraints with $W(s,x,y,\widehat{s})$ to obtain valid inequalities in the space of $(Q_{X|S}, Q_{\widehat{S}|Y}, W)$. Clearly, these constraints are implied by the constraints of SC. To obtain the LP relaxation, the constraint $W(s,x,y,\widehat{s}) = Q_{X|S}(x|s)Q_{\widehat{S}|Y}(\widehat{s}|y)$ for all $s,x,y,\widehat{s}$ is dropped and the new valid inequalities in the space of $(Q_{X|S}, Q_{\widehat{S}|Y}, W)$ are added to the original constraints of SC. Thus, the following is our LP relaxation.

$$
\begin{array}{lll}
\text{LP} & \min\limits_{Q_{X|S},Q_{\widehat{S}|Y},W} \sum\limits_z \kappa(z)P_S(s)P_{Y|X}(y|x)W(z) & \\
\text{s.t} & \sum_x Q_{X|S}(x|s) = 1 : \gamma^a(s) & \forall s \\
& \sum_{\widehat{s}} Q_{\widehat{S}|Y}(\widehat{s}|y) = 1 : \gamma^b(y) & \forall y \\
& \sum_x W(z) - Q_{\widehat{S}|Y}(\widehat{s}|y) = 0 : \lambda_{\mathsf{s}}(s,\widehat{s},y) & \forall s,\widehat{s},y \\
& \sum_{\widehat{s}} W(z) - Q_{X|S}(x|s) = 0 : \lambda_{\mathsf{c}}(x,s,y) & \forall x,s,y \\
& Q_{X|S}(x|s) + Q_{\widehat{S}|Y}(\widehat{s}|y) - W(z) \leq 1 : \mu(z) & \forall z \\
& Q_{X|S}(x|s) \geq 0 & \forall s,x \\
& Q_{\widehat{S}|Y}(\widehat{s}|y) \geq 0 & \forall \widehat{s},y \\
& W(z) \geq 0 & \forall z.
\end{array}
$$

Here $\gamma^a, \gamma^b, \lambda_{\mathsf{s}}$,$\lambda_{\mathsf{c}}$ and μ are Lagrange multipliers corresponding to the respective constraints. The following property holds:

Lemma 2.1: The extreme points of the feasible region of SC are included in the extreme points of the feasible region of LP.

Employing the Lagrange multipliers $\gamma^a, \gamma^b, \lambda_{\mathsf{s}}, \lambda_{\mathsf{c}}$ and μ corresponding to the constraints in LP, we obtain the dual program denoted by DP.

$$
\begin{array}{lll}
\text{DP} & \max\limits_{\gamma^a,\gamma^b,\lambda^a,\lambda^b,\mu} \sum\limits_s \gamma^a(s) + \sum\limits_y \gamma^b(y) - \sum\limits_z \mu(z) & \\
\text{s.t} & \gamma^a(s) - \sum_y \lambda^b(x,s,y) - \sum_{\widehat{s},y} \mu(z) \leq 0 & \forall x,s \\
& \gamma^b(y) - \sum_s \lambda^a(s,\widehat{s},y) - \sum_{x,s} \mu(z) \leq 0 & \forall \widehat{s},y \\
& \lambda^a(s,\widehat{s},y) + \lambda^b(x,s,y) + \mu(z) \leq \vartheta(z) & \forall z \\
& \mu(z) \geq 0 & \forall z
\end{array}
$$

where $\vartheta(z) = \kappa(z)P_S(s)P_{Y|X}(y|x)$ for all z. Henceforth, we call these Lagrange multipliers as the 'dual variables'. Notice that these variables are restricted in their domain and in the case of μ, also their range. Specifically, $\gamma^a : \mathcal{S} \to \mathbb{R}$, $\gamma^b : \mathcal{Y} \to \mathbb{R}$, $\lambda_{\mathsf{c}} : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $\lambda_{\mathsf{s}} : \mathcal{S} \times \widehat{\mathcal{S}} \times \mathcal{Y} \to \mathbb{R}$ and $\mu : \mathcal{Z} \to \mathbb{R}_+$, whereby these variables are functions on *subspaces* of $\mathcal{Z}$. The following lemma formalizes our framework for obtaining lower bounds on SC.

Lemma 2.2: The objective value of any feasible point of DP is a lower bound on the optimal value of SC, i.e., if $\big(\gamma^a(s), \gamma^b(y), \lambda_{\mathsf{s}}(s,\widehat{s},y), \lambda_{\mathsf{c}}(x,s,y), \mu(s,x,y,\widehat{s})\big)_{z\in\mathcal{Z}}$ is feasible for DP, then

$$
\begin{aligned}
\mathrm{OPT(SC)} &\geq \mathrm{OPT(LP)} = \mathrm{OPT(DP)} \\
&\geq \sum_s \gamma^a(s) + \sum_y \gamma^b(y) - \sum_{s,x,y,\widehat{s}} \mu(s,x,y,\widehat{s}). \quad (2)
\end{aligned}
$$

LP can be further relaxed by removing the inequality, $-1 + Q_{X|S}(x|s) + Q_{\widehat{S}|Y}(\widehat{s}|y) - W(x,s,\widehat{s},y) \leq 0$, for all $s,x,y,\widehat{s}$ from the problem LP to result in a weaker relaxation, LP$'$. The dual of LP$'$ is same as DP with the dual variable $\mu(s,x,y,\widehat{s})$ set identically to 0. In DP$'$, it is optimal to take $\gamma^a(s)$ and $\gamma^b(y)$ such that (D1) and (D2) hold with equality. Thus, the optimal value of DP$'$ with $\Sigma(z)$ as the RHS of (D3) evaluates to, $\mathrm{OPT(DP}', \Sigma(z)) =$

$$
\begin{aligned}
&\max_{\lambda_{\mathsf{s}},\lambda_{\mathsf{c}}} \Bigg\{ \sum_s \inf_x \sum_y \lambda_{\mathsf{c}}(x,s,y) + \sum_y \inf_{\widehat{s}} \sum_s \lambda_{\mathsf{s}}(s,\widehat{s},y) \Bigg\} \\
&\text{s.t} \qquad \lambda_{\mathsf{c}}(x,s,y) + \lambda_{\mathsf{s}}(s,\widehat{s},y) \leq \Sigma(z). \quad (3)
\end{aligned}
$$

Applying Lemma 2.2 with LP$'$ as relaxation then yields that any choice of functions $(\lambda_{\mathsf{s}}, \lambda_{\mathsf{c}})|_{z\in\mathcal{Z}}$ satisfying constraint (D3), yields the following lower bound on OPT(SC),

$$
\begin{aligned}
\mathrm{OPT(SC)} &\geq \mathrm{OPT(LP}') = \mathrm{OPT(DP}') \\
&\geq \sum_s \inf_x \sum_y \lambda_{\mathsf{c}}(x,s,y) + \sum_y \inf_{\widehat{s}} \sum_s \lambda_{\mathsf{s}}(s,\widehat{s},y). \quad (4)
\end{aligned}
$$

Note that λ_{c} is a function of the source signal s and the channel input and outputs x,y. Consequently, we refer to this as a *channel flow*. On the other hand, λ_{s} is independent of channel input x, but is a function of source and destination signals $s,\widehat{s}$. Hence, λ_{s} can be referred to as a *source flow*. Thus, in the joint source-channel coding problem with $S \to X \to Y \to \widehat{S}$ forming a Markov chain, λ_{c} and λ_{s} constitute a channel flow and a source flow, respectively, through this path. Moreover, from DP$'$, it is evident that the choice of these flows is dictated by the bottleneck on the " error density", $\Sigma(z) = \kappa(s,x,y,\widehat{s})P_S(s)P_{Y|X}(y|x)$, imposed by the constraint (D3). As such, evaluating OPT(DP$'$) corresponds to finding the optimal packing of source and channel flows satisfying the error density bottleneck.

III. LP-Based Finite Blocklength Converses in Point-to-Point Information Theory

In this section, we present our new LP-based finite blocklength converses for three main coding problems in point-to-point information theory- lossy joint source-channel coding, lossy source coding and channel coding. The results can be found in our paper [8].

A. Lossy Joint Source-Channel Coding

For the lossy joint source-channel coding problem, we consider problem SC with S having the distribution P_S and channel conditional probability distribution $P_{Y|X}$. The cost function is given as $\kappa(s,x,y,\widehat{s}) \equiv \mathbb{I}\{d(s,\widehat{s}) > \mathbf{d}\}$, where $d : \mathcal{S} \times \widehat{\mathcal{S}} \to [0,+\infty]$ represents the distortion function and $\mathbf{d} \in [0,\infty)$ is the distortion level. The objective is to obtain a lower bound on the minimum value of $\mathbb{E}[\mathbb{I}\{d(S,\widehat{S}) > \mathbf{d}\}] = \mathbb{P}[d(S,\widehat{S}) > \mathbf{d}]$ (which is called the excess distortion probability) achieved by a joint source-channel code (f,g). We will use DP' to derive a lower bound on this problem.

Theorem 3.1: (DP improves Generalized Kostina-Verdú bound) Consider problem SC with S having distribution P_S and channel conditional probability distribution given by $P_{Y|X}$. Let $\kappa(s,x,y,\widehat{s}) \equiv \mathbb{I}\{d(s,\widehat{s}) > \mathbf{d}\}$ be the loss function where $d : \mathcal{S} \times \widehat{\mathcal{S}} \to [0,+\infty]$ is a distortion function and $\mathbf{d} \in [0,\infty)$ is the distortion level. Then, there exists a feasible point of the problem DP' with an objective value which improves on the Kostina-Verdú converse [3, Theorem 3]. Specifically, for any code, we have the following lower bound on excess distortion probability,

$$\begin{aligned}
&\mathbb{E}[\mathbb{I}\{d(S,\widehat{S}) > \mathbf{d}\}] \geq \mathrm{OPT}(\mathrm{SC}) \geq \mathrm{OPT}(\mathrm{LP}') = \mathrm{OPT}(\mathrm{DP}') \\
&\geq \max_{\gamma,T}\Bigg\{\sup_{\bar{Y},V}\mathbb{E}\Bigg[\inf_{x}\Bigg\{\mathbb{P}[j_S(S,\mathbf{d}) - i_{X;\bar{Y}|V}(x;Y|V) \geq \gamma|S] \\
&+ \exp(j_S(S,\mathbf{d}) - \gamma)\sum_{y\in\mathcal{Y}}\sum_{t=1}^{T} P_{V|X}(t|x)P_{\bar{Y}|V}(y|t)\times \\
&\qquad \mathbb{I}\{j_S(S,\mathbf{d}) - i_{X;\bar{Y}|V}(x;y|t) < \gamma\}\Bigg\}\Bigg] \\
&- T\sup_{\widehat{s}}\mathbb{E}\Bigg[\exp(j_S(S,\mathbf{d}) - \gamma)\mathbb{I}\{d(S,\widehat{s}) \leq \mathbf{d}\}\Bigg]\Bigg\}, \qquad (5)
\end{aligned}$$

where $j_S(S,\mathbf{d})$ is the $\mathbf{d}$-tilted information defined in [3], T is a positive integer, V is a random variable that takes values on $\{1,2,\ldots,T\}$,

$$i_{X;\bar{Y}|V}(x;y|t) = \log\frac{P_{Y|X=x,V=t}}{P_{\bar{Y}|V=t}}(y), \qquad (6)$$

$\mathbb{P}$ is generated by $P_{V|X=x}P_{Y|X=x,V}$. Note that the expectation $\mathbb{E}$ on the RHS of (5) is with respect to P_S.

The presence of an additional non-negative term corresponding to

$$\mathbb{I}\{j_S(S,\mathbf{d}) - i_{X;\bar{Y}|V}(x;y|t) < \gamma\}$$

in (5) and the fact that

$$T\sup_{\widehat{s}}\mathbb{E}\Bigg[\exp(j_S(S,\mathbf{d}) - \gamma)\mathbb{I}\{d(S,\widehat{s}) \leq \mathbf{d}\}\Bigg] < T\exp(-\gamma)$$

which follows from the property of $\mathbf{d}$-tilted information, result in the improvement on the Kostina-Verdú converse in [3, Theorem 3].

Finally, if $T = 1$ in the converse in Theorem 3.1, we an improvement on the Kostina-Verdú converse in [3, Theorem 1].

B. Lossy Source Coding

For the finite blocklength lossy source coding problem, consider problem SC with $\mathcal{X} = \mathcal{Y} = \{1,\ldots M\}$ and $P_{Y|X}(y|x) \equiv \mathbb{I}\{y = x\}$. We consider the probability of excess distortion as the loss criterion with $\kappa(s,x,y,\widehat{s}) \equiv \mathbb{I}\{d(s,\widehat{s}) > \mathbf{d}\}$, where $d : \mathcal{S} \times \widehat{\mathcal{S}} \to [0,\infty]$ is the distortion function and $\mathbf{d} \geq 0$ is the distortion level.

Particularizing the improved lossy joint source-channel converse in (5) with $T = 1$ to lossy source coding, results in the following converse which improves on the Kostina-Verdú converse in [2, Theorem 7].

Corollary 3.2: (Lossy Source Coding – Further Improvement) Consider the setting of Theorem 3.1 with $\mathcal{X} = \mathcal{Y} = \{1,\ldots,M\}$, $M \in \mathbb{N}$ and channel conditional distribution $P_{Y|X}(y|x) \equiv \mathbb{I}\{x = y\}$. Then, for any code, the following lower bound follows from (5),

$$\begin{aligned}
&\mathbb{E}[\mathbb{I}\{d(S,\widehat{S}) > \mathbf{d}\}] \geq \mathrm{OPT}(\mathrm{SC}) \geq \mathrm{OPT}(\mathrm{DP}') \\
&\quad \geq \sup_{\gamma}\Bigg\{\mathbb{P}[j_S(S,\mathbf{d}) \geq \gamma + \log M] \\
&+ \frac{1}{M}\sum_{s} P_S(s)\exp(j_S(s,\mathbf{d}) - \gamma)\mathbb{I}\{j_S(s,\mathbf{d}) < \log M + \gamma\} \\
&- \sup_{\widehat{s}}\exp(-\gamma)\sum_{s} P_S(s)\exp(j_S(s,\mathbf{d}))\mathbb{I}\{d(s,\widehat{s}) \leq \mathbf{d}\}\Bigg\}, \qquad (7)
\end{aligned}$$

which improves on the Kostina-Verdú converse in [2, Theorem 7].

We now proceed to obtain a new LP based metaconverse for lossy source coding, which recovers the hypothesis testing based converse in [2, Theorem 8], our improvement on the tilted information based Kostina and Verdú converse in (7) and even the converse of Roy and Lars [9].

Theorem 3.3 (Metaconverse for Lossy Source Coding): Consider problem SC. For any code,

$$\begin{aligned}
&\mathbb{E}[\mathbb{I}\{d(S,\widehat{S}) > \mathbf{d}\}] \geq \mathrm{OPT}(\mathrm{SC}) \geq \mathrm{OPT}(\mathrm{LP}') \\
&\geq \sup_{0\leq\phi(s)\leq P_S(s)}\Bigg\{\sum_{s}\phi(s) - M\max_{\widehat{s}}\sum_{s}\phi(s)\mathbb{I}\{d(s,\widehat{s}) \leq \mathbf{d}\}\Bigg\}, \qquad (8)
\end{aligned}$$

where the supremum is over all functions $\phi : \mathcal{S} \to [0,1]$ such that $0 \leq \phi(s) \leq P_S(s)$ for all $s \in \mathcal{S}$.

Proof : To obtain the required bound, consider the following values of dual variables,

$$\begin{aligned}
\lambda_{\mathsf{c}}(x,s,y) &\equiv \mathbb{I}\{y = x\}\phi(s), \quad \lambda_{\mathsf{s}}(s,\widehat{s},y) \equiv -\phi(s)\mathbb{I}\{d(s,\widehat{s}) \leq \mathbf{d}\} \\
\gamma^a(s) &\equiv \sum_{y}\lambda^b(x,s,y), \quad \gamma^b(y) \equiv \inf_{\widehat{s}}\sum_{s}\lambda^a(s,\widehat{s},y).
\end{aligned}$$

Feasibility of these variables with respect to dual constraints can be easily verified. Consequently, the required lower bound follows from Lemma 2.2 by taking supremum over ϕ such that $0 \leq \phi(s) \leq P_S(s)\ \forall s \in \mathcal{S}$. ■

In particular, choosing $\phi(s) = \min\{P_S(s), z(s)\}$ in (8) where

$z : \mathcal{S} \to [0, \infty)$, and taking supremum over such z, we get the following bound,

$$\mathbb{E}[\mathbb{I}\{d(S, \widehat{S}) > \mathbf{d}\}] \geq \mathrm{OPT(SC)} \geq \mathrm{OPT(LP')}$$
$$\geq \sup_{z \geq 0} \Bigg\{ \sum_s \min\{P_S(s), z(s)\}$$
$$- M \max_{\widehat{s}} \sum_s \min\{P_S(s), z(s)\} \mathbb{I}\{d(s, \widehat{s}) \leq \mathbf{d}\} \Bigg\}. \quad (9)$$

We now show that our LP based meta-converse implies all other known converses.

1) It is straightforward to see that (9) in fact implies the hypothesis testing based converse [2, Theorem 8]. To see this, take $z_s = \beta Q(s)$ in (8) and subsequently, take supremum over $\beta \geq 0$ and $Q \in \mathcal{P}(\mathcal{S})$. Consequently, we have,

$$\mathbb{E}[\mathbb{I}\{d(S, \widehat{S}) > \mathbf{d}\}] \geq \mathrm{OPT(SC)} \geq \mathrm{OPT(LP)}$$
$$\geq \sup_{\beta \geq 0} \sup_{Q_S \in \mathcal{P}(\mathcal{S})} \Bigg\{ \sum_s \min\{P_S(s), \beta Q_S(s)\}$$
$$- M \sup_{\widehat{s}} \sum_s \min\{P_S(s), \beta Q_S(s)\} \mathbb{I}\{d(s, \widehat{s}) \leq \mathbf{d}\} \Bigg\}$$
$$\geq \sup_{\beta \geq 0} \sup_{Q_S \in \mathcal{P}(\mathcal{S})} \Bigg\{ \sum_s \min\{P_S(s), \beta Q_S(s)\} - M^* \Bigg\},$$
$$= \sup_{Q_S \in \mathcal{P}(\mathcal{S})} \alpha_{M^*}(P_S, Q_S)$$

which is the hypothesis testing based converse in [2, Theorem 8], where $M^* = \sup_{\widehat{s}} \sum_s \beta Q_S(s) \mathbb{I}\{d(s, \widehat{s}) \leq \mathbf{d}\}$.

2) In particular, taking $z_s = \beta \mu(s)$ in (9), where μ is some positive, finite measure (not necessarily probability measure) on $\mathcal{S}$, we have,

$$\sup_{\beta, \mu} \sum_s \min\{P_S(s), \beta \mu(s)\} - M \sup_{\widehat{s}} \beta \sum_s \mu(s) \mathbb{I}\{d(s, \widehat{s}) \leq \mathbf{d}\}$$
$$= \alpha_{M \sup_{\widehat{s}} \mu[d(S, \widehat{s}) \leq \mathbf{d}]}(P_S, \mu),$$

which is the converse obtained by Kostina in [10, Theorem 3].

3) To obtain converse (7) from our metaconverse (9), take $z_s = P_S(s) \frac{1}{M} \exp(j_S(s, \mathbf{d}) - \gamma)$ and subsequently lower bound $-M \sup_{\widehat{s}} \sum_s \min\{P_S(s), z_s\} \mathbb{I}\{d(s, \widehat{s}) \leq \mathbf{d}\}$ with $-M \sup_{\widehat{s}} \sum_s z_s \mathbb{I}\{d(s, \widehat{s}) \leq \mathbf{d}\}$, which results in converse 7.

4) To obtain Palzer et al. converse [9], take $z_s = P_S(s) \mathbb{I}\{j_S(s, \mathbf{d}) \geq \beta\}$ in (9) and take supremum over $\beta \geq 0$. Notice that in this case, $\min\{z_s, P_S(s)\} = z_s$ since $z_s = P_S(s) \mathbb{I}\{j_S(s, \mathbf{d}) \geq \beta\} \leq P_S(s)$.

C. Channel Coding

For finite blocklength channel coding problem, consider problem SC with $\mathcal{S} = \widehat{\mathcal{S}} = \{1, \ldots, M\}$, $P_S(s) \equiv \frac{1}{M}$ and $\kappa(s, x, y, \widehat{s}) \equiv \mathbb{I}\{s \neq \widehat{s}\}$ such that $\mathbb{E}[\kappa(S, X, Y, \widehat{S})] = \mathbb{P}[S \neq \widehat{S}]$ gives the average probability of error.

Particularizing the improved finite blocklength converse for lossy joint source-channel coding in (5) to channel coding by considering $\mathcal{S} = \widehat{\mathcal{S}} = \{1, \ldots, M\}$, $P_S(s) \equiv \frac{1}{M}$, $d(s, \widehat{s}) \equiv \mathbb{I}\{s \neq \widehat{s}\}$ and $\mathbf{d} = 0$, results in the following improvement on the converse of Wolfowitz citewolfowitz1968notes.

Corollary 3.4 (DP′ improves on Wolfowitz's Converse): Consider the setting of Corollary 3.1 with $\mathcal{S} = \widehat{\mathcal{S}} = \{1, 2, \ldots, M\}$, $M \in \mathbb{N}$, $P_S(s) \equiv \frac{1}{M}$. Let $d(s, \widehat{s}) \equiv \mathbb{I}\{s \neq \widehat{s}\}$ and $\mathbf{d} = 0$. Then, for any code, the following lower bound on the minimum error probability holds,

$$\mathbb{E}[\mathbb{I}\{S \neq \widehat{S}\}] \geq \mathrm{OPT(SC)} \geq \mathrm{OPT(DP')}$$
$$\geq \sup_{\gamma} \Bigg\{ \sup_{P_{\bar{Y}}} \inf_x \Bigg[\mathbb{P}[i_{X;\bar{Y}}(x; Y) \leq \log M - \gamma]$$
$$+ M \exp(-\gamma) \sum_{y \in \mathcal{Y}} P_{\bar{Y}}(y) \mathbb{I}\{i_{X;\bar{Y}}(x; y) > \log M - \gamma\} \Bigg]$$
$$- \exp(-\gamma) \Bigg\}. \quad (10)$$

Further, the above converse implies a strong converse for the discrete memoryless channels.

The improvement on Wolfowitz's converse is on account of the additional non-negative term corresponding to $\mathbb{I}\{i_{X;\bar{Y}}(x; y) > \log M - \gamma\}$.

The following theorem shows that the optimal value of DP′ in fact coincides with the metaconverse of Polyanskiy-Poor-Verdú (PPV) [1].

Theorem 3.5 (DP implies PPV Metaconverse [1]): Consider problem SC with $\mathcal{S} = \widehat{\mathcal{S}} = \{1, \ldots, M\}$, $\kappa(s, x, y, \widehat{s}) \equiv \mathbb{I}\{s \neq \widehat{s}\}$ and S is uniformly distributed on $\mathcal{S}$. Consequently, for any code, the following bound holds,

$$\mathbb{E}[\mathbb{I}\{S \neq \widehat{S}\}] \geq \mathrm{OPT(SC)} \geq \mathrm{OPT(LP')} = \mathrm{OPT(DP')}$$
$$\overset{(a)}{=} \sup_{z_y \geq 0} \Bigg\{ \inf_x \sum_y \min\{P_{Y|X=x}(y), z_y\} - \frac{\sum_y z_y}{M} \Bigg\},$$
$$= \inf_{P_X} \sup_{P_{\bar{Y}}} \alpha_{\frac{1}{M}}(P_X P_{Y|X}, P_X P_{\bar{Y}}), \quad (11)$$

which is in fact the PPV metaconverse.

Proof : To obtain the converse in (a), consider the following values of dual variables,

$$\lambda_{\mathsf{c}}(x, s, y) \equiv \frac{1}{M} \min\{P_{Y|X=x}(y), z_y\},$$
$$\lambda_{\mathsf{s}}(s, \widehat{s}, y) \equiv -\frac{z_y}{M} \mathbb{I}\{s = \widehat{s}\},$$

and γ^a, γ^b such that (D1), (D2) hold with equality. It is easy to verify that the above dual variables satisfy the dual constraints. Consequently, the dual cost becomes,

$$\sum_s \gamma^a(s) + \sum_y \gamma^b(y) = \inf_x \sum_y \min\{P_{Y|X=x}(y), z_y\} - \frac{\sum_y z_y}{M}.$$

which upon taking supremum over $z_y \geq 0$ yields the required bound. The equivalence of $\mathrm{OPT(DP')}$ to the above bound follows from the result of Matthews [11]. ■

Moreover, our improvement on the converse of Wolfowitz

obtained in (10) is in fact equivalent to the Polyanskiy-Poor-Verdú lower bound (11). To see this, note that while the supremum is over $z : \mathcal{Y} \to [0, \infty)$ in (11), setting $z_y = P_{\bar{Y}}(y) M \exp(-\gamma)$ in (11) and replacing maximum over z_y with supremum over $\gamma > 0$ and $P_{\bar{Y}}$ results in (10).

IV. LP-based Framework to Averaged and Compound Channels

This section concerns with finite blocklength converses for two special classes of channels - *averaged* and *compound* channels under the *maximum probability of error* criterion- an error criterion not addressed in the point-to-point setting. In Section III-B, we considered the point-to-point channel coding problem SC. With $P_S(s) \equiv \frac{1}{M}$, the average probability of error achieved by a code (f, g) for this channel is then defined as,

$$\epsilon_{\text{avg}} = \sum_{s,x,y,\widehat{s}} \frac{1}{M} \mathbb{I}\{x = f(s)\} P_{Y|X}(y|x) \mathbb{I}\{\widehat{s} = g(y)\} \mathbb{I}\{s \neq \widehat{s}\},$$

and maximum probability of error is defined as,

$$\epsilon_{\max} = \max_{s \in \mathcal{S}} \sum_{x,y,\widehat{s}} \mathbb{I}\{x = f(s)\} P_{Y|X}(y|x) \mathbb{I}\{\widehat{s} = g(y)\} \mathbb{I}\{s \neq \widehat{s}\}.$$

Interestingly, for discrete memoryless point-to-point channels, a strong converse is known to hold under both average and maximum probability of error criterion. In fact, the strong converse is implied by the converse of Wolfowitz. Consequently, the capacity of discrete memoryless channels remains unchanged under the probability of error criterion. However, this property does not hold true for any channel in general.

Averaged channel, a term coined by Ahslwede in [12] and also known by *mixed channel* is a special class of channel for which a weak converse is known to hold under the maximum probability of error and not a strong converse. On the other hand, for the compound channels, a weak and strong converse holds under maximum probability of error criterion and not under the average probability of error criterion. While in the point-to-point channel coding setting the error criterion considered was immaterial, it plays a crucial role in defining the capacity of averaged and compound channels. Consequently, the central theme of this section is to develop the LP relaxation based framework to account for the maximum probability of error criterion and derive new finite blocklength converses such that they imply a weak converse for averaged channel and a strong converse for compound channel.

Let Θ be a finite set of indices and θ be a random variable taking values in Θ. For an averaged channel, the channel conditional probability distribution $\bar{P}_{Y|X}$ is given as,

$$\bar{P}_{Y|X}(y|x) \equiv \sum_{\theta \in \Theta} q(\theta) P_{Y|X,\theta}(y|x, \theta), \tag{12}$$

which is obtained by averaging the channel conditional probability distributions, $P_{Y|X,\theta}, \theta \in \Theta$ according to the known distribution $q \in \mathcal{P}(\Theta)$. Without loss of generality, we take $q(\theta) > 0$ for all $\theta \in \Theta$. The maximum probability of error incurred by the averaged channel on employing the code (f, g) is given by,

$$\kappa_{\text{avg}} \equiv \max_{s \in \mathcal{S}} \sum_{y} \bar{P}_{Y|X}(y|f(s)) \mathbb{I}\{g(y) \neq s\}. \tag{13}$$

On the other hand, a *compound channel* represents the finite collection of channels, $\{P_{Y|X,\theta}, \theta \in \Theta\}$. In a compound channel, for transmitting a signal $x = (x_1, \ldots, x_n)$, a channel is selected arbitrarily from the collection and is held fixed throughout the transmission of the block. Let κ_{com} represent the maximum probability of error achieved by a code (f, g) when transmitting through the compound channel. Precisely,

$$\kappa_{\text{com}} \equiv \max_{\theta \in \Theta, s \in \mathcal{S}} \sum_{y} P_{Y|X,\theta}(y|f(s), \theta) \mathbb{I}\{g(y) \neq s\}. \tag{14}$$

While under average probability of error criterion, the finite blocklength channel coding problem SC could be equivalently posed as an optimization problem over randomized codes, this luxury is lost when dealing with maximum probability of error criterion. This is because the maximum probability of error criterion calls for finding the maximum error incurred over all messages $S \in \{1, \ldots, M\}$ for a given deterministic code (f, g). Consequently, randomizing the codes so as to pose the fininte blocklength channel coding problem as a nonconvex optimization problem over joint probability distributions results in a relaxed version of the original problem. Employing the LP relaxation approach on this nonconvex, already relaxed version of the original problem, further relaxes the problem. Hence, under maximum probability of error criterion, we effectively consider two-level relaxations, which imply two avenues for incurring losses.

The resulting dual program becomes,

$$\boxed{\begin{array}{lll} \text{DP}_\Delta & \max\limits_{\gamma^a, \gamma^b, \lambda^a, \lambda^b, \mu} \sum\limits_{s} \gamma^a(s) + \sum\limits_{y} \gamma^b(y) & \\ & \gamma^a(s) - \sum_y \lambda^b(x, s, y) \leq 0 & \forall x, s \\ \text{s.t} & \gamma^b(y) - \sum_s \lambda^a(s, \widehat{s}, y) \leq 0 & \forall \widehat{s}, y \\ & \lambda^a(s, \widehat{s}, y) + \lambda^b(x, s, y) \leq \Upsilon_\Delta(z) & \forall s, \widehat{s}, y \\ & \sum_\delta \mu(\delta) \leq 1, \quad \mu(\delta) \geq 0 & \forall \delta, \end{array}}$$

where DP_Δ represents the dual program corresponding to $\Delta \in \{\text{avg}, \text{com}\}$, avg representing averaged channel and com representing the compound channel. Moreover, $\delta = s$ if $\Delta = \text{avg}$ and $\delta = (s, \theta)$ if $\Delta = \text{com}$ and

$$\Upsilon_{\text{avg}}(z) \equiv \mu(s) \bar{P}_{Y|X}(y|x) \mathbb{I}\{s \neq \widehat{s}\},$$

$$\Upsilon_{\text{com}}(z) \equiv \sum_{\theta} \mu(s, \theta) P_{Y|X,\theta}(y|x, \theta) \mathbb{I}\{s \neq \widehat{s}\}.$$

We refer the readers to [13] for more details.

Employing the LP-based framework results in the following new finite blocklength converses for averaged and compound channels.

Theorem 4.1: Consider the channel coding of averaged channel with $\mathcal{S} = \widehat{\mathcal{S}} = \{1, \dots M\}$ and $\bar{P}_{Y|X}$ as in (12). Consequently, for any code,

$$\begin{aligned}\kappa_{\text{avg}} &\geq \text{OPT}(\text{DP}_{\text{avg}}) \\ &\geq \sup_{\bar{Y},\beta>0} \Bigg\{ \min_{x\in\mathcal{X}} \sum_{\theta\in\Theta} q(\theta) \Bigg[\mathbb{P}[i_{X;\bar{Y}|\theta}(x;Y|\theta) \leq \log\beta] \\ &+ \beta \sum_{y} P_{\bar{Y}|\theta}(y|\theta) \mathbb{I}\{i_{X;\bar{Y}|\theta}(x;y|\theta) > \log\beta\} \Bigg] - \frac{\beta}{M} \Bigg\}, \quad (15)\end{aligned}$$

where $i_{X;\bar{Y}|\theta}(x;y|\theta) = \log \frac{P_{Y|X=x,\theta}}{P_{\bar{Y}|\theta}}(y)$, $P_{\bar{Y}|\theta}$ is an arbitrary probability distribution on $\mathcal{Y}$ and $\mathbb{P}$ in the RHS of (15) is with respect to $P_{Y|X=x,\theta}$.

Let $\mathcal{X} = \mathcal{A}^n$, $\mathcal{Y} = \mathcal{B}^n$, where $\mathcal{A}$ and $\mathcal{B}$ represent the finite channel input alphabet and output alphabet, respectively. A Discrete Memoryless Averaged Channel (DMAC) is given by the distribution $\bar{P}_{Y|X}$ as defined in (12) with

$$P_{Y|X,\theta}(y|x,\theta) \equiv \prod_{i=1}^{n} P_{\mathbf{Y}|\mathbf{X},\theta}(y_i|x_i,\theta) \quad \forall \theta \in \Theta, \qquad (16)$$

where $P_{\mathbf{Y}|\mathbf{X},\theta} \in \mathcal{P}(\mathcal{B}|\mathcal{A},\theta)$. A 'weak capacity' for the DMAC was obtained by Ahlswede [12] as

$$\mathcal{C}_{\text{avg}} = \max_{P_X \in \mathcal{P}(\mathcal{A})} \min_{\theta\in\Theta} I_{P_X}(\mathbf{X};\mathbf{Y}|\theta),$$

where $I_{P_X}(\mathbf{X};\mathbf{Y}|\theta)$ is the mutual information of $(\mathbf{X},\mathbf{Y}) \sim P_X P_{\mathbf{Y}|\mathbf{X},\theta}$.

Consider the DMAC with $\mathcal{S} = \widehat{\mathcal{S}} = \{1, \dots, \exp(nR)\}$. A code (f,g) for this DMAC is called a $(\exp(nR), n)$ code and $\kappa_{\text{avg}}^{(n)}$ represents the maximum probability of error achieved by this code. Then, the following result follows from (15).

Theorem 4.2: For the DMAC $\bar{P}_{Y|X}$ with $P_{Y|X,\theta}$ as given in (16), if there exists a sequence of $(\exp(nR), n)$ codes such that $\kappa_{\text{avg}}^{(n)}$ goes to zero as $n \to \infty$, then the converse in (15) implies that $R < \mathcal{C}_{\text{avg}}$.

Proof : See [13]. ■

By constructing a feasible point of DP_{com}, we obtain the following new lower bound on κ_{com}.

Theorem 4.3: Consider the channel coding of compound channels with $\mathcal{S} = \widehat{\mathcal{S}} = \{1, \dots M\}$. For any code,

$$\begin{aligned}\kappa_{\text{com}} &\geq \text{OPT}(\text{DP}_{\text{com}}) \\ &\geq \sup_{\bar{Y},\beta>0,\mu(s,\theta)\geq 0} \Bigg\{ \sum_{s} \min_{x\in\mathcal{X}} \sum_{\theta\in\Theta} \mu(s,\theta) \Bigg[\mathbb{P}[i_{X;\bar{Y}|\theta}(x;Y|\theta) \leq \log\beta] \\ &+ \beta \sum_{y} P_{\bar{Y}|\theta}(y|\theta) \mathbb{I}\{i_{X;\bar{Y}\theta}(x;y|\theta) > \log\beta\} \Bigg] \\ &\quad - \beta \sum_{y} \sup_{\widehat{s}} \sum_{\theta} P_{\bar{Y}|\theta}(y|\theta) \mu(\widehat{s},\theta) \Bigg\}, \qquad (17)\end{aligned}$$

where $\sum_{s,\theta} \mu(s,\theta) \leq 1$, $\mathbb{P}$ is with respect to $P_{Y|X=x,\theta}$.

Consider a Discrete Memoryless Compound Channel (DMCC) $\{P_{Y|X,\theta}, \theta \in \Theta\}$ with component channels $P_{Y|X,\theta}$ are as given in (16). The capacity of this DMCC is given as $\mathcal{C}_{\text{com}} = \max_{P_X\in\mathcal{P}(\mathcal{A})} \min_{\theta\in\Theta} I_{P_X}(\mathbf{X};\mathbf{Y}|\theta)$. Notice that $\mathcal{C}_{\text{com}} = \mathcal{C}_{\text{avg}}$. However, it is not clear whether the converse in (17) directly implies a strong converse. Let $\kappa_{\text{com}}^{(n)}$ be the maximum probability of error achieved by a code (f,g). To show the strong converse for DMCC, i.e., to show that if $R > \mathcal{C}_{\text{com}}$, then $\lim_{n\to\infty} \kappa_{\text{com}}^{(n)} = 1$, we resort to an auxiliary problem with restricted spaces of $\mathcal{S}$ and $\mathcal{X}$ as explained in the following theorem.

Theorem 4.4: Let $\kappa_{\text{com}}^{(n)}$ be the maximum probability of error achieved by a code (f,g) for the DMCC. Let P_n be a type on $\mathcal{A}$. Consider problem CC_{com} with $\mathcal{X}' = \{x \in f(\mathcal{S}) \mid x \text{ has type } P_n\}$ and $\widehat{\mathcal{S}}' = \mathcal{S}' = f^{-1}(\mathcal{X}')$ with $|\mathcal{S}'| = M'$ and $\mathcal{Y} = \mathcal{B}^n$. Let $\bar{\kappa}_{\text{com}}^{(n)}$ represent the maximum probability of error achieved by the code (f,g) restricted to this setup. Consequently, for any code, $\kappa_{\text{com}}^{(n)} \geq \bar{\kappa}_{\text{com}}^{(n)}$ and

$$\begin{aligned}\bar{\kappa}_{\text{com}}^{(n)} &\geq \text{OPT}(\text{DP}_{\text{com}}) \\ &\geq \sup_{\bar{Y},\theta',\gamma} \Bigg\{ \min_{x\in\mathcal{X}'} \mathbb{P}[i_{X;\bar{Y}|\theta}(x;Y|\theta') \leq \log M - O(\log n) - \gamma] \\ &- \exp(-\gamma) \Bigg\}. \qquad (18)\end{aligned}$$

To show the strong converse for the DMCC, let $R = \mathcal{C}_{\text{com}} + \eta$, $\eta > 0$, lower bound (18) by setting $\theta' = \bar{\theta}$ such that $I_{P_n}(\mathbf{X};\mathbf{Y}|\bar{\theta}) \leq \mathcal{C}_{\text{com}}$. Notice that $\bar{\kappa}_{\text{com}}^{(n)}$, the type P_n and consequently $\bar{\theta}$ depend on n. Further, take, $P_{\bar{Y}|\theta}(y|\theta) \equiv (P_n P_{\mathbf{Y}|\mathbf{X},\bar{\theta}})^{\times n}(y)$ and $\gamma = n\frac{\eta}{4}$. It is then easy to verify the strong converse.

V. LP-Based Framework to Multi-terminal Source Coding Problems

A. Finite Blocklength Slepian-Wolf Coding

Consider the finite blocklength Slepian-Wolf distributed lossless source coding problem posed as the following optimization problem,

$$\begin{array}{ll} \text{SW} & \min_{f_1,f_2,g} \ \mathbb{E}[\mathbb{I}\{(S_1,S_2) \neq (\widehat{S}_1,\widehat{S}_2)\}] \\ & X_1 = f_1(S_1), \\ \text{s.t} & X_2 = f_2(S_2), \\ & (\widehat{S}_1,\widehat{S}_2) = g(Y_1,Y_2), \end{array}$$

where $S_1, S_2, X_1, X_2, Y_1, Y_2, \widehat{S}_1, \widehat{S}_2$ are discrete random variables taking values in fixed, finite spaces $\mathcal{S}_1, \mathcal{S}_2, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2, \widehat{\mathcal{S}}_1, \widehat{\mathcal{S}}_2$, respectively. Here, S_1 and S_2 represent the two *correlated* sources distributed according to a known joint probability distribution P_{S_1,S_2}. The source signals are *seperately* encoded by functions $f_1 : \mathcal{S}_1 \to \mathcal{X}_1$ and $f_2 : \mathcal{S}_2 \to \mathcal{X}_2$ to produce signals $X_1 = f_1(S_1)$ and $X_2 = f_2(S_2)$, respectively. The encoded signals are sent through a deterministic channel with conditional distribution $P_{Y_1,Y_2|X_1,X_2} = \mathbb{I}\{(Y_1,Y_2) = (X_1,X_2)\}$ to get the signal (Y_1,Y_2). (Y_1,Y_2) is then *jointly* decoded by $g : \mathcal{Y}_1 \times \mathcal{Y}_2 \to \widehat{\mathcal{S}}_1 \times \widehat{\mathcal{S}}_2$ to obtain the output signal $(\widehat{S}_1,\widehat{S}_2) = g(Y_1,Y_2)$. For the finite blocklength Slepian-Wolf coding problem, we note that spaces $\mathcal{S}_1 = \widehat{\mathcal{S}}_1$, $\mathcal{S}_2 = \widehat{\mathcal{S}}_2$, $\mathcal{X}_1 = \mathcal{Y}_1 = \{1, \dots, M_1\}$ and $\mathcal{X}_2 = \mathcal{Y}_2 = \{1, \dots, M_2\}$,

$M_1, M_2 \in \mathbb{N}$. An error in transmission occurs when $(S_1, S_2) \neq (\widehat{S}_1, \widehat{S}_2)$. Hence, the objective of the finite blocklength Slepian-Wolf coding problem SW is to minimize the probability of error over all *codes*, i.e. over all encoder-decoder functions (f_1, f_2, g). Our objective is to obtain a lower bound on the optimal value of SW and we achieve this via a linear programming relaxation of SW.

Extending the LP relaxation approach for the point-to-point setting to network setting results in the following dual program corresponding to SW, DPSW. Here, $\Pi(z) \equiv \mathbb{I}\{(s_1, s_2) \neq (\widehat{s}_1, \widehat{s}_2)\} P_{S_1,S_2}(s_1, s_2) \mathbb{I}\{(y_1, y_2) = (x_1, x_2)\}$. Let $\Theta := (\lambda_{\mathsf{s}}^{(1|2)}, \lambda_{\mathsf{s}}^{(2|1)}, \lambda_{\mathsf{c}}, \gamma^a, \gamma^b, \gamma^c, \mu_{\mathsf{s}}^{(1)}, \mu_{\mathsf{c}}^{(1)}, \mu_{\mathsf{s}}^{(2)}, \mu_{\mathsf{c}}^{(2)}, \mu_{\mathsf{c}}^{(1|2)}, \mu_{\mathsf{c}}^{(2|1)})$ represent the collection of all these Lagrange multipliers or dual variables.

As is evident, construction of a feasible point of DPSW is challenging and probably cumbersome at first glance. Another hindrance is the difficulty in interpreting these variables so as to develop any intuitions on construction of these variables.

Consequently, we present a systematic method to synthesize new converses for problem SW from the converses for the following sub-problems of SW, (a) lossless source coding of jointly encoded correlated sources (S_1, S_2), (b) lossless source coding of S_1 with perfect side-information of S_2 available at the decoder, and (c) lossless source coding of S_2 with perfect side-information of S_1 at the decoder.

1) Lossless Coding of Jointly Encoded Correlated Sources (S_1, S_2): In this sub-problem of Slepian-Wolf coding problem, the correlated sources S_1, S_2 are *jointly encoded* according to $f : \mathcal{S}_1 \times \mathcal{S}_2 \to \mathcal{X}_1 \times \mathcal{X}_2$ to get (X_1, X_2). (X_1, X_2) is sent through the channel $P_{Y_1,Y_2|X_1,X_2} = \mathbb{I}\{(Y_1, Y_2) = (X_1, X_2)\}$ to get (Y_1, Y_2) which is then decoded according to $g : \mathcal{Y}_1 \times \mathcal{Y}_2 \to \widehat{\mathcal{S}}_1 \times \widehat{\mathcal{S}}_2$. The objective, as for SW problem, is to losslessly recover (S_1, S_2) at the destination.

It is easy to see that the above joint encoding problem is in fact equivalent to the point-to-point lossless source coding problem SC in Section III-B with $S := (S_1, S_2)$, $X := (X_1, X_2)$, $Y := (Y_1, Y_2)$, $\widehat{S} := (\widehat{S}_1, \widehat{S}_2)$ and $d(S, \widehat{S}) = \mathbb{I}\{S \neq \widehat{S}\}$ with $\mathbf{d} = 0$. Consequently, we have the following generalized version of DP′ for lossless source coding problem,

$$\begin{aligned}
\text{DPJE} \quad & \max_{\widehat{\gamma}^a, \widehat{\gamma}^b, \widehat{\lambda}_{\mathsf{s}}, \widehat{\lambda}_{\mathsf{c}}} \sum_{s_1,s_2} \widehat{\gamma}^a(s_1, s_2) + \sum_{y_1,y_2} \widehat{\gamma}^b(y_1, y_2) \\
\text{s.t} \quad & \widehat{\gamma}^a(s_1, s_2) - \textstyle\sum_{y_1,y_2} \widehat{\lambda}_{\mathsf{c}}(s_1, s_2, x_1, x_2, y_1, y_2) \leq 0 \quad \forall x_1, x_2, s_1, s_2 \quad \text{(A1)} \\
& \widehat{\gamma}^b(y_1, y_2) - \textstyle\sum_{s_1,s_2} \widehat{\lambda}_{\mathsf{s}}(s_1, s_2, \widehat{s}_1, \widehat{s}_2, y_1, y_2) \leq 0 \quad \forall \widehat{s}_1, \widehat{s}_2, y_1, y_2 \quad \text{(A2)} \\
& \widehat{\lambda}_{\mathsf{s}}(s_1, s_2, \widehat{s}_1, \widehat{s}_2, y_1, y_2) + \widehat{\lambda}_{\mathsf{c}}(s_1, s_2, x_1, x_2, y_1, y_2) \leq \Upsilon(\widetilde{z}) \quad \forall \widetilde{z} \quad \text{(A3)}
\end{aligned}$$

where $\widetilde{z} := (s_1, s_2, x_1, x_2, y_1, y_2, \widehat{s}_1, \widehat{s}_2)$, $\Upsilon(\widetilde{z}) = \mathbb{I}\{(s_1, s_2) \neq (\widehat{s}_1, \widehat{s}_2)\} P_{S_1,S_2}(s_1, s_2) \mathbb{I}\{(y_1, y_2) = (x_1, x_2)\}$ for all $\widetilde{z}$.

The metaconverse for lossless source coding problem which follows from Theorem 3.3 then readily implies the following corollary.

Corollary 5.1 (Metaconverse for Jointly Encoded Sources): Consider problem SC with $S := (S_1, S_2)$, $X := (X_1, X_2)$, $Y := (Y_1, Y_2)$, $\widehat{S} := (\widehat{S}_1, \widehat{S}_2)$ and $d(S, \widehat{S}) = \mathbb{I}\{S \neq \widehat{S}\}$ with $\mathbf{d} = 0$. Consequently for any code, the following metaconverse holds,

$$\begin{aligned}
&\mathbb{E}[\mathbb{I}\{(S_1, S_2) \neq (\widehat{S}_1, \widehat{S}_2)\}] \geq \text{OPT(DPJE)} \\
&\geq \sup_{0 \leq \widehat{\phi}(s_1,s_2) \leq P_{S_1,S_2}(s_1,s_2)} \left\{ \sum_{s_1,s_2} \widehat{\phi}(s_1, s_2) - M_1 M_2 \max_{\widehat{s}_1,\widehat{s}_2} \widehat{\phi}(\widehat{s}_1, \widehat{s}_2) \right\},
\end{aligned} \tag{19}$$

where the supremum is over $\widehat{\phi} : \mathcal{S}_1 \times \mathcal{S}_2 \to [0, 1]$ such that $0 \leq \widehat{\phi}(s_1, s_2) \leq P_{S_1,S_2}(s_1, s_2)$ for all $s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2$.

Proof : To obtain the required converse, consider the following choice for the flows in DPJE,

$$\begin{aligned}
\widehat{\lambda}_{\mathsf{c}}(x_1, x_2, s_1, s_2, y_1, y_2) &\equiv \mathbb{I}\{(y_1, y_2) = (x_1, x_2)\} \widehat{\phi}(s_1, s_2) \\
\widehat{\lambda}_{\mathsf{s}}(s_1, s_2, \widehat{s}_1, \widehat{s}_2, y_1, y_2) &\equiv -\widehat{\phi}(s_1, s_2) \mathbb{I}\{(s_1, s_2) = (\widehat{s}_1, \widehat{s}_2)\},
\end{aligned} \tag{20}$$

and $\widehat{\gamma}^a$, $\widehat{\gamma}^b$ chosen such that (A1), (A2) hold with equality. ■

2) Lossless Source Coding of S_1 with S_2 as the Side-Information: In this section, we consider the following sub-problem of Slepian-Wolf coding : S_1 is to be recovered losslessly at the destination with S_2 available as side-information at the decoder . Towards this, S_1 is encoded according to $f_1 : \mathcal{S}_1 \to \mathcal{X}_1$ to get X_1, which is transmitted through the channel $P_{Y_1|X_1} = \mathbb{I}\{Y_1 = X_1\}$ to get Y_1. S_2 is the side information available at the decoder which decodes according to $g : \mathcal{S}_2 \times \mathcal{Y}_1 \to \widehat{\mathcal{S}}_1$ to get $\widehat{S}_1$.

Employing the LP relaxation approach results in the following dual program,

$$\begin{aligned}
\text{DPSI}_{1|2} \quad & \max_{\bar{\gamma}^a, \bar{\gamma}^b, \bar{\lambda}_{\mathsf{s}}^{(1|2)}, \bar{\lambda}_{\mathsf{c}}^{(1|2)}} \sum_{s_1} \bar{\gamma}^a(s_1) + \sum_{y_1,s_2} \bar{\gamma}^b(y_1, s_2) \\
\text{s.t} \quad & \bar{\gamma}^a(s_1) - \textstyle\sum_{y_1,s_2} \bar{\lambda}_{\mathsf{c}}^{(1|2)}(s_1, s_2, x_1, y_1) \leq 0 \quad \forall x_1, s_1 \quad \text{(B1)} \\
& \bar{\gamma}^b(s_2, y_1) - \textstyle\sum_{s_1} \bar{\lambda}_{\mathsf{s}}^{(1|2)}(s_1, s_2, \widehat{s}_1, y_1) \leq 0 \quad \forall s_2, \widehat{s}_1, y_1 \quad \text{(B2)} \\
& \bar{\lambda}_{\mathsf{s}}^{(1|2)}(s_1, s_2, \widehat{s}_1, y_1) + \bar{\lambda}_{\mathsf{c}}^{(1|2)}(s_1, s_2, x_1, y_1) \leq \Psi(\bar{z}) \quad \forall \bar{z} \quad \text{(B3)}
\end{aligned}$$

where $\bar{z} := (x_1, s_1, s_2, \widehat{s}_1, y_1)$ and $\Psi(\bar{z}) = P_{S_1,S_2}(s_1, s_2) \mathbb{I}\{s_1 \neq \widehat{s}_1\} \mathbb{I}\{y_1 = x_1\}$.

Theorem 5.2: Consider the problem $\text{SID}_{1|2}$. For any code, the following lower bound holds,

$$\begin{aligned}
&\mathbb{E}[\mathbb{I}\{S_1 \neq \widehat{S}_1\}] \geq \text{OPT}(\text{DPSI}_{1|2}) \geq \sup_{0 \leq \phi^{(1|2)} \leq P_{S_1,S_2}} \left\{ \sum_{s_1,s_2} \phi^{(1|2)}(s_1, s_2) \right. \\
&\left. - M_1 \sum_{s_2} \max_{\widehat{s}_1} \phi^{(1|2)}(\widehat{s}_1, s_2) \right\},
\end{aligned} \tag{21}$$

where the supremum is over $\phi^{(1|2)} : \mathcal{S}_1 \times \mathcal{S}_2 \to [0, 1]$ such that $\phi^{(1|2)}(s_1, s_2) \leq P_{S_1,S_2}(s_1, s_2)$ for all $s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2$.

$$
\begin{array}{lrll}
\text{DPSWmax}_{\Theta} & \sum_{s_1}\gamma^a(s_1)+\sum_{s_2}\gamma^b(s_2)+\sum_{y_1,y_2}\gamma^c(y_1,y_2) & & \\
& \gamma^a(s_1)-\sum_{y_1,y_2}\mu_{\mathsf{c}}^{(1)}(s_1,x_1,y_1,y_2)-\sum_{s_2}\mu_{\mathsf{c}}^{(1|2)}(x_1,s_1,s_2)\le 0 & \forall x_1,s_1 & (D1)\\
& \gamma^b(s_2)-\sum_{y_1,y_2}\mu_{\mathsf{c}}^{(2)}(s_2,x_2,y_1,y_2)-\sum_{s_1}\mu_{\mathsf{c}}^{(2|1)}(x_2,s_1,s_2)\le 0 & \forall x_2,s_2 & (D2)\\
& \gamma^c(y_1,y_2)-\sum_{s_2}\mu_{\mathsf{s}}^{(2)}(s_2,\widehat{s}_1,\widehat{s}_2,y_1,y_2)-\sum_{s_1}\mu_{\mathsf{s}}^{(1)}(s_1,\widehat{s}_1,\widehat{s}_2,y_1,y_2)\le 0 & \forall \widehat{s}_1,\widehat{s}_2,y_1,y_2 & (D3)\\
\text{subject to} & \lambda_{\mathsf{s}}^{(1|2)}(s_1,s_2,x_2,y_1,y_2,\widehat{s}_1,\widehat{s}_2)+\lambda_{\mathsf{s}}^{(2|1)}(s_1,s_2,x_1,y_1,y_2,\widehat{s}_1,\widehat{s}_2)+\lambda_{\mathsf{c}}(s_1,s_2,x_1,x_2,y_1,y_2)\le\Pi(z) & \forall z & (D4)\\
& \mu_{\mathsf{s}}^{(2)}(s_2,\widehat{s}_1,\widehat{s}_2,y_1,y_2)+\mu_{\mathsf{c}}^{(2)}(s_2,x_2,y_1,y_2)-\sum_{s_1}\lambda_{\mathsf{s}}^{(1|2)}(s_1,s_2,x_2,y_1,y_2,\widehat{s}_1,\widehat{s}_2)\le 0 & \forall s_2,x_2,y_1,y_2,\widehat{s}_1,\widehat{s}_2 & (D5)\\
& \mu_{\mathsf{s}}^{(1)}(s_1,\widehat{s}_1,\widehat{s}_2,y_1,y_2)+\mu_{\mathsf{c}}^{(1)}(s_1,x_1,y_1,y_2)-\sum_{s_2}\lambda_{\mathsf{s}}^{(2|1)}(s_1,s_2,x_1,y_1,y_2,\widehat{s}_1,\widehat{s}_2)\le 0 & \forall s_1,x_1,y_1,y_2,\widehat{s}_1,\widehat{s}_2 & (D6)\\
& \mu_{\mathsf{c}}^{(2|1)}(x_2,s_1,s_2)+\mu_{\mathsf{c}}^{(1|2)}(x_1,s_1,s_2)-\sum_{y_1,y_2}\lambda_{\mathsf{c}}(s_1,s_2,x_1,x_2,y_1,y_2)\le 0 & \forall x_1,x_2,s_1,s_2 & (D7)
\end{array}
$$

Proof : To obtain the required converse, we consider the following values for the source flow and channel flow,

$$
\begin{aligned}
\bar{\lambda}_{\mathsf{c}}^{(1|2)}(s_1,s_2,x_1,y_1)&=\mathbb{I}\{x_1=y_1\}\phi^{(1|2)}(s_1,s_2),\\
\bar{\lambda}_{\mathsf{s}}^{(1|2)}(s_1,s_2,\widehat{s}_1,y_1)&=-\phi^{(1|2)}(s_1,s_2)\mathbb{I}\{s_1=\widehat{s}_1\},
\end{aligned}
\tag{22}
$$

and $\gamma^a(s_1)$, $\gamma^b(s_2,y_1)$ chosen such that (B1), (B2) hold with equality. ■

3) Lossless Source Coding of S_2 with S_1 as the Side-Information: Analogous to $\mathrm{SID}_{1|2}$, the finite blocklength lossless source coding problem of S_2 with S_1 as the side information yields the following dual program employing the LP relaxation approach,

$$
\begin{array}{lll}
\mathrm{DPSI}_{2|1} & \max\limits_{\tilde{\gamma}^a,\tilde{\gamma}^b,\tilde{\lambda}_{\mathsf{s}}^{(2|1)},\tilde{\lambda}_{\mathsf{c}}^{(2|1)}} \sum_{s_2}\tilde{\gamma}^a(s_2)+\sum_{y_2,s_1}\tilde{\gamma}^b(y_2,s_1) & \\
& \tilde{\gamma}^a(s_2)-\sum_{y_2,s_1}\tilde{\lambda}_{\mathsf{c}}^{(2|1)}(s_1,s_2,x_2,y_2)\le 0 & \\
& \qquad \forall x_2,s_2 \quad (C1) & \\
\text{s.t} & \tilde{\gamma}^b(s_1,y_2)-\sum_{s_2}\tilde{\lambda}_{\mathsf{s}}^{(2|1)}(s_1,s_2,\widehat{s}_2,y_2)\le 0 & \\
& \qquad \forall s_1,\widehat{s}_2,y_2 \quad (C2) & \\
& \tilde{\lambda}_{\mathsf{s}}^{(2|1)}(s_1,s_2,\widehat{s}_2,y_2)+\tilde{\lambda}_{\mathsf{c}}^{(2|1)}(s_1,s_2,x_2,y_2)\le\Delta(z') & \\
& \qquad \forall z' \quad (C3), &
\end{array}
$$

where $z' := (s_1,s_2,x_2,y_2,\widehat{s}_2)$, and $\Delta(z') = P_{S_1,S_2}(s_1,s_2)\mathbb{I}\{y_2=x_2\}\mathbb{I}\{s_2\neq\widehat{s}_2\}$. Taking $\tilde{\gamma}^a(s_2)$ and $\tilde{\gamma}^b(s_1,y_2)$ such that (C1) and (C2) hold with equality, the following choice of source flow and channel flow results in a converse similar to the one in Theorem 5.2 for the problem $\mathrm{SID}_{2|1}$.

$$
\begin{aligned}
\tilde{\lambda}_{\mathsf{c}}^{(2|1)}(s_1,s_2,x_2,y_2)&=\phi^{(2|1)}(s_1,s_2)\mathbb{I}\{x_2=y_2\}\\
\tilde{\lambda}_{\mathsf{s}}^{(2|1)}(s_1,s_2,\widehat{s}_2,y_2)&=-\phi^{(2|1)}(s_1,s_2)\mathbb{I}\{s_2=\widehat{s}_2\},
\end{aligned}
\tag{23}
$$

where $\phi^{(2|1)} : \mathcal{S}_1\times\mathcal{S}_2\to[0,1]$ is such that $0\le\phi^{(2|1)}(s_1,s_2)\le P_{S_1,S_2}(s_1,s_2)$ for all $s_1\in\mathcal{S}_1, s_2\in\mathcal{S}_2$.

Theorem 5.3: Consider the problem $\mathrm{SID}_{2|1}$. Consequently, for any code, the following lower bound holds,

$$
\begin{aligned}
&\mathbb{E}[\mathbb{I}\{S_2\neq\widehat{S}_2\}]\ge\mathrm{OPT}(\mathrm{SID}_{2|1})\ge\mathrm{OPT}(\mathrm{DPSI}_{2|1})\\
&\ge\sup_{0\le\phi^{(2|1)}(s_1,s_2)\le P_{S_1,S_2}(s_1,s_2)}\Bigg\{\sum_{s_1,s_2}\phi^{(2|1)}(s_1,s_2)\\
&\qquad -M_2\sum_{s_1}\max_{\widehat{s}_2}\phi^{(2|1)}(s_1,\widehat{s}_2)\Bigg\}.
\end{aligned}
\tag{24}
$$

4) Synthesizing Converses for SW from Point-to-Point Duals: In this section, we present a systematic synthesis of finite blocklength converses for the Slepian-Wolf coding problem from the source and channel flows of the dual programs DPJE, $\mathrm{DPSI}_{1|2}$ and $\mathrm{DPSI}_{2|1}$.

Theorem 5.4: Let $(\bar{\gamma}^a,\bar{\gamma}^b,\bar{\lambda}_{\mathsf{s}}^{(1|2)},\bar{\lambda}_{\mathsf{c}}^{(1|2)})\in\mathrm{FEA}(\mathrm{DPSI}_{1|2})$, $(\tilde{\gamma}^a,\tilde{\gamma}^b,\tilde{\lambda}_{\mathsf{s}}^{(2|1)},\tilde{\lambda}_{\mathsf{c}}^{(2|1)})\in\mathrm{FEA}(\mathrm{DPSI}_{2|1})$ and $(\widehat{\gamma}^a,\widehat{\gamma}^b,\widehat{\lambda}_{\mathsf{s}},\widehat{\lambda}_{\mathsf{c}})\in\mathrm{FEA}(\mathrm{DPJE})$. Then, any choice of values for the variables of DPSW satisfying the following equations is feasible for DPSW.

$$
\begin{aligned}
&\lambda_{\mathsf{s}}^{(1|2)}(s_1,s_2,x_2,y_1,y_2,\widehat{s}_1,\widehat{s}_2)=\Bigg[\bar{\lambda}_{\mathsf{s}}^{(1|2)}(s_1,s_2,\widehat{s}_1,y_1)\times\\
&\mathbb{I}\{x_2=y_2\}+\alpha\widehat{\lambda}_{\mathsf{s}}(s_1,s_2,\widehat{s}_1,\widehat{s}_2,y_1,y_2)\Bigg]\mathbb{I}\{(s_1,s_2)=(\widehat{s}_1,\widehat{s}_2)\},\\
&\lambda_{\mathsf{s}}^{(2|1)}(s_1,s_2,x_1,y_1,y_2,\widehat{s}_1,\widehat{s}_2)=\Bigg[\tilde{\lambda}_{\mathsf{s}}^{(2|1)}(s_1,s_2,\widehat{s}_2,y_2)\times\\
&\mathbb{I}\{x_1=y_1\}+(1-\alpha)\widehat{\lambda}_{\mathsf{s}}(s_1,s_2,\widehat{s}_1,\widehat{s}_2,y_1,y_2)\Bigg]\times\\
&\qquad\qquad\mathbb{I}\{(s_1,s_2)=(\widehat{s}_1,\widehat{s}_2)\},\\
&\lambda_{\mathsf{c}}(s_1,s_2,x_1,x_2,y_1,y_2)=\min\Bigg\{P(s_1,s_2)\mathbb{I}\{(y_1,y_2)=(x_1,x_2)\},\\
&\qquad\widehat{\lambda}_{\mathsf{c}}(s_1,s_2,x_1,x_2,y_1,y_2)+\bar{\lambda}_{\mathsf{c}}^{(1|2)}(s_1,s_2,x_1,y_1)\mathbb{I}\{x_2=y_2\}\\
&\qquad+\tilde{\lambda}_{\mathsf{c}}^{(2|1)}(s_1,s_2,x_2,y_2)\mathbb{I}\{x_1=y_1\}\Bigg\},
\end{aligned}
\tag{25}
$$

$$
\begin{aligned}
&\mu_{\mathsf{c}}^{(2)}(s_2,x_2,y_1,y_2)\le\Bigg[\bar{\gamma}^b(y_1,s_2)-\sum_{s_1\neq\widehat{s}_1}\bar{\lambda}_{\mathsf{s}}^{(1|2)}(s_1,s_2,\widehat{s}_1,y_2)\Bigg]\\
&\qquad\qquad\times\mathbb{I}\{x_2=y_2\}\mathbb{I}\{s_2=\widehat{s}_2\},\\
&\mu_{\mathsf{s}}^{(2)}(s_2,\widehat{s}_1,\widehat{s}_2,y_1,y_2)=\alpha\widehat{\lambda}_{\mathsf{s}}(\widehat{s}_1,s_2,\widehat{s}_1,\widehat{s}_2,y_1,y_2)\mathbb{I}\{s_2=\widehat{s}_2\},\\
&\mu_{\mathsf{c}}^{(1)}(s_1,x_1,y_1,y_2)\le\Bigg[\tilde{\gamma}^b(s_1,y_2)-\sum_{s_2\neq\widehat{s}_2}\tilde{\lambda}_{\mathsf{s}}^{(2|1)}(s_2,s_2,\widehat{s}_2,y_2)\Bigg]\\
&\qquad\qquad\times\mathbb{I}\{x_1=y_1\}\mathbb{I}\{s_1=\widehat{s}_1\},\\
&\mu_{\mathsf{s}}^{(1)}(s_1,\widehat{s}_1,\widehat{s}_2,y_1,y_2)=(1-\alpha)\widehat{\lambda}_{\mathsf{s}}(s_1,\widehat{s}_2,\widehat{s}_1,\widehat{s}_2,y_1,y_2)\\
&\qquad\qquad\times\mathbb{I}\{s_1=\widehat{s}_1\},\\
&\mu_{\mathsf{c}}^{(2|1)}(x_2,s_1,s_2)+\mu_{\mathsf{c}}^{(1|2)}(x_1,s_1,s_2)\\
&\qquad\le\sum_{y_1,y_2}\lambda_{\mathsf{c}}(s_1,s_2,x_1,x_2,y_1,y_2),
\end{aligned}
$$

where $\alpha\in(0,1)$ and $\gamma^a(s_1),\gamma^b(s_2)$ and $\gamma^c(y_1,y_2)$ are chosen such that (D1), (D2) and (D3) hold with equality.

Proof : See [14]. ■

Thanks to Theorem 5.4, to obtain finite blocklength converses for Slepian-Wolf coding, it suffices to consider the simpler point-to-point source-coding problems and construct good feasible points for them. In particular, considering those feasible points of $\mathrm{DPSI}_{1|2}$, $\mathrm{DPSI}_{2|1}$ and DPJE which yield the meta-converses in (21), (24) and (19) for the corresponding point-to-point sub-problems and subsequently employing Theorem 5.4, we obtain the following new finite blocklength converse for SW.

Theorem 5.5 (Metaconverse for Slepian-Wolf Coding): Consider the problem SW. Consequently, for any code, the following bound holds:

$$\begin{aligned}&\mathbb{E}[\mathbb{I}\{(S_1,S_2)\neq(\widehat{S}_1,\widehat{S}_2)\}]\geq \mathrm{OPT}(\mathrm{SW})\geq \mathrm{OPT}(\mathrm{DPSW})\geq\\ &\sup_{\widehat{\phi},\phi^{(1|2)},\phi^{(2|1)}}\Bigg\{\sum_{s_1,s_2}\min\{P_{S_1,S_2}(s_1,s_2),\widehat{\phi}(s_1,s_2)+\phi^{(1|2)}(s_1,s_2)\\ &+\phi^{(2|1)}(s_1,s_2)\}-M_1M_2\max_{\widehat{s}_1,\widehat{s}_2}\widehat{\phi}(\widehat{s}_1,\widehat{s}_2)\\ &-M_2\sum_{s_1}\max_{\widehat{s}_2}\phi^{(2|1)}(s_1,\widehat{s}_2)-M_1\sum_{s_2}\max_{\widehat{s}_1}\phi^{(1|2)}(\widehat{s}_1,s_2)\Bigg\},\end{aligned}\tag{26}$$

where the supremum is over $\widehat{\phi},\phi^{(1|2)},\phi^{(2|1)}:\mathcal{S}_1\times\mathcal{S}_2\to[0,1]$ such that $0\leq\widehat{\phi}(s_1,s_2),\phi^{(1|2)}(s_1,s_2),\phi^{(2|1)}(s_1,s_2)\leq P_{S_1,S_2}(s_1,s_2)$ for all $s_1\in\mathcal{S}_1,s_2\in\mathcal{S}_2$.

In particular, choosing

$$\begin{aligned}\widehat{\phi}(s_1,s_2)&=\min\{P_{S_1,S_2}(s_1,s_2),\eta_1(s_1,s_2)\}\\ \phi^{(1|2)}(s_1,s_2)&=\min\{P_{S_1,S_2}(s_1,s_2),\eta_2(s_1,s_2)\}\\ \phi^{(2|1)}(s_1,s_2)&=\min\{P_{S_1,S_2}(s_1,s_2),\eta_3(s_1,s_2)\}\end{aligned}$$

in (26), results in the following improvement on the converse of Miyake and Kanaya [15],

$$\begin{aligned}&\mathbb{E}[\mathbb{I}_{\{(S_1,S_2)\neq(\widehat{S}_1,\widehat{S}_2)\}}]\geq \mathrm{OPT}(\mathrm{SC})\geq \mathrm{OPT}(\mathrm{DP})\geq\\ &\sup_{\beta>0}\Bigg\{\mathbb{P}\Bigg[h_{S_1,S_2}(S_1,S_2)\geq\log M_1M_2+\beta\ \text{ or }\ h_{S_1|S_2}(S_1|S_2)\\ &\geq\log M_1+\beta\text{ or }\ h_{S_2|S_1}(S_2|S_1)\geq\log M_2+\beta\Bigg]+\\ &\sum_{s_1,s_2}\max\Bigg\{\frac{\exp(-\beta)}{M_1M_2},\frac{\exp(-\beta)}{M_1}P_{S_2}(s_2),\frac{\exp(-\beta)}{M_2}P_{S_1}(s_1)\Bigg\}\\ &\times\mathbb{I}\Bigg\{P_{S_1|S_2}(s_1|s_2)>\frac{\exp(-\beta)}{M_1},P_{S_2|S_1}(s_2|s_1)>\frac{\exp(-\beta)}{M_2},\\ &\quad P_{S_1,S_2}(s_1,s_2)>\frac{\exp(-\beta)}{M_1M_2}\Bigg\}-3\exp(-\beta)\Bigg\},\end{aligned}\tag{27}$$

where $h_{A|B}(a|b)\equiv-\log P_{A|B}(a|b)$ is the conditional entropy density and $h_{A,B}(a,b)\equiv-\log P_{A,B}(a,b)$ is the joint entropy density.

VI. Conclusion

A unified Linear Programming based framework is presented to derive new finite blocklength converses for coding problems in information theory. The framework is shown to be applicable for any loss criteria and even extendable to network settings.

References

[1] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, 2010.

[2] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 58, no. 6, pp. 3309–3338, 2012.

[3] ——, "Lossy joint source-channel coding in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 59, no. 5, pp. 2545–2575, 2013.

[4] L. Zhou, V. Y. Tan, and M. Motani, "Second-order and moderate deviations asymptotics for successive refinement," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2896–2921, 2017.

[5] A. A. Kulkarni and T. P. Coleman, "An optimizer's approach to stochastic control problems with nonclassical information structures," in *Proceedings of the IEEE Conference on Decision and Control*, 2012.

[6] S. T. Jose and A. A. Kulkarni, "A linear programming based channel coding strong converse for the BSC and BEC," in *Proceedings of the 23rd National Conference on Communications, 2017, (To appear)*, 2017.

[7] M. Conforti, G. Cornuéjols, and G. Zambelli, *Integer programming*. Springer, 2014, vol. 271.

[8] S. T. Jose and A. A. Kulkarni, "Linear programming based converses for finite-blocklength joint-source channel coding," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7066–7094, 2017.

[9] L. Palzer and R. Timo, "A converse for lossy source coding in the finite blocklength regime," 2016.

[10] V. Kostina, "Data compression with low distortion and finite blocklength," *IEEE Transactions on Information Theory*, 2017.

[11] W. Matthews, "A linear program for the finite block length converse of polyanskiy–poor–verdú via nonsignaling codes," *IEEE Transactions on Information Theory*, vol. 58, no. 12, pp. 7036–7044, 2012.

[12] R. Ahlswede, "The weak capacity of averaged channels," *Probability Theory and Related Fields*, vol. 11, no. 1, pp. 61–73, 1968.

[13] S. T. Jose and A. A. Kulkarni, "Linear programming based converses for some network-like problems," in *to appear in the Proceedings of Information Theory Workshop*, 2017.

[14] ——, "Improved finite blocklength converses for Slepian-Wolf coding via linear programming," *under review, IEEE Transactions on Information Theory*, 2018.

[15] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 78, no. 9, pp. 1063–1070, 1995.